

Statistical assessment on determining local presence of rare bat species

Kathryn M. Irvine¹, Katharine M. Banner², Christian Stratton², William M. Ford³, Brian E. Reichert⁴

¹*U.S. Geological Survey, Northern Rocky Mountain Science Center, Bozeman, Montana, USA*

²*Montana State University, Department of Mathematical Sciences, Bozeman, Montana, USA*

³*U.S. Geological Survey, Virginia Cooperative Fish and Wildlife Unit, Blacksburg, Virginia, USA*

⁴*U.S. Geological Survey, Fort Collins Science Center, Fort Collins, Colorado, USA*

Corresponding author: Kathryn M. Irvine, email: kirvine@usgs.gov

March 11, 2022

Running Header: Assessing presence for project clearance

Open Research Statement: All data, novel simulation code, and novel model code used in this manuscript are archived https://github.com/kbanner14/SuppS2_IrvineEtAl-ecosphere.

This repository will be made publicly available through Zenodo if this work is published.

Code to fit the Wright et al. (2020) count detection is available at Wright et al. (2019).

Abstract Surveying cryptic, sparsely distributed taxa using autonomous recording units, although cost-effective, provides imperfect knowledge about species presence. Summertime bat acoustic surveys in North America exemplify the challenges with characterizing sources of uncertainty: observation error, inability to census populations, and natural stochastic variation. Statistical uncertainty, if not considered thoroughly, hampers determining rare species presence accurately and/or estimating rangewide status and trends with suitable precision. Bat acoustic data are processed using an automated workflow in which proprietary or open-source algorithms assign a species label to each recorded high-frequency echolocation sequence. A false-negative occurs, if a species is actually present but not recorded and/or all recordings from the species are of such poor quality that a correct species identity cannot be assigned to any observation. False positives for a focal species are a direct result of the presence and incorrect identification of a recording from another species. We compare four analytical approaches in terms of parameter estimation and their resulting (in)correct decisions regarding species presence or absence using realistic data-generating scenarios for bat acoustic data within a simulation study. The current standard for deciding species presence or absence uses a multinomial likelihood ratio test p-value (MLE-metric) that accounts for known species misidentifications, but not imperfect detection and only returns a binary outcome (evidence of presence or not). We found that the MLE-metric had estimated median correct decisions less than 60% for presence and greater than 85% for absence. Alternatively, a multi-species count detection model was equivalent to or better than the MLE-metric for correct claims of rare species presence or absence using the posterior probability a species was present at a site and, importantly, provided unbiased estimates of relative activity and probability of occurrence, creating opportunities for reducing

posterior uncertainty through inclusion of meaningful covariates. Single-species occupancy models with and without false-positive detections removed were insufficient for determining local presence because of substantially biased occurrence and detection probabilities. We propose solutions to potential barriers for integrating local, short-term and rangewide, long-term acoustic surveys within a cohesive statistical framework that facilitates determining local species presence with uncertainty concurrent with estimating species-environment relationships.

Keywords: Acoustic Survey, Autonomous Recording Units, Bayesian Hierarchical Model, Count Detection Model, False Positives, Imperfect Detection, North American Bat Monitoring Program, Occupancy Modeling, Sampling Design

1. INTRODUCTION

Decision-making in the face of uncertainty is a fundamental challenge for natural resource managers (Polasky et al., 2011). Recommendations related to a species conservation status under the U.S. Endangered Species Act (ESA 1973, as amended) are typically assessed at a rangewide extent using long-term monitoring datasets, when possible (e.g., Smith et al. 2018). Status assessments consider a species’ persistence on the landscape under plausible scenarios for future conditions related to their known stressors (e.g., in the case of bats, impacts of climate change, energy development, forest management, and disease; Frick et al., 2020; Hoyt et al., 2021). Ideally, spatially extensive and longitudinal datasets are collated to fuel statistical models for estimating with uncertainty whether populations are increasing, decreasing, or stable over time under hypothetical scenarios (Erickson et al., 2014; Cheng et al., 2021). Once a species is deemed “threatened or endangered” under ESA, approval for a habitat alteration often requires a determination of species presence or absence within a potentially affected area to minimize possible “take” of vulnerable species (USFWS, 2020). Data collection to inform the permitting process, typically, relies on surveys over a shorter time period within a given project area (e.g., for *Myotis sodalis* and *M. septentrionalis* see USFWS, 2020). Our work is motivated by considering whether a common statistical inference framework is available for addressing both rangewide status assessments and local decisions regarding rare bat species presence using data collected by autonomous recording units (ARUs)—acoustic surveys.

The observational unit for bat acoustic surveys is a high-frequency recording that can be

visualized using a spectrogram (Aodha et al., 2018). Many species can be recorded during a nightly interval at a specific deployment location (a “visit”). A nuance of bat acoustic data is that the sheer volume of echolocation sequences recorded during a visit necessitates an automated workflow for managing the recordings (Aodha et al., 2018). Recordings are processed using proprietary classification software or open-source machine learning algorithms to assign or label an echolocation sequence to a particular species (Acevedo et al., 2009; Sugai et al., 2018). Consequently, cross-species misidentifications—the wrong species is assigned to a recording—can be introduced during the automated classification process. Misclassifications at the recording level, if not accounted for, can propagate into visit-level false-positive and false-negative detections that can bias resulting statistical inferences (Clement et al., 2014; Chambert, Waddle, Miller, Walls and Nichols, 2018; Banner et al., 2018).

To minimize the potential for a software program to “confuse” echolocation sequences among bats in the genus *Myotis* and *Lasiurus borealis*, the “maximum likelihood” metric was developed based on misclassification rates calculated from high-frequency recordings with known species identities (the “MLE”-metric; Britzke et al., 2002). The MLE-metric is the current standard supported by the U.S. Fish and Wildlife Service (USFWS) for determining local presence for two rare bat species, *M. septentrionalis* and *M. sodalis*. To ensure consistency in application among project areas, the USFWS independently approves automated bat acoustic identification software programs (i.e., auto-classifier) for acoustic survey work (USFWS, 2020). The MLE-metric returned by approved software is then used to make a decision about species presence at a site. For regulatory purposes, the USFWS considers *M. septentrionalis* and *M. sodalis* present at a local site, if one or more survey nights (visits) returns an MLE-metric less than 0.05. To date, no work has explored the performance

of the MLE-metric for determining local bat presence using a statistical simulation study. The benefit of our simulation investigation is that the true state (species present or not) is known and available for investigating how often the MLE-metric approach arrives at a correct decision under different data-generating scenarios for bat acoustic data.

We consider whether Bayesian occupancy models are a viable alternative to the MLE-metric for determining local presence of rare bat species within the same simulation study. Often biologists summarize the nightly recordings to detection/non-detection (one or zero) for a single focal species and apply occupancy models for statistical analyses (Rodhouse et al., 2012). Standard occupancy models account for imperfect detection (MacKenzie et al., 2002) at the visit level, which for bat ARU data, arise because a species is actually present but not recorded, and/or all recordings from the species are of such poor quality that a correct species identity cannot be assigned to any observation (Banner et al., 2018). A visit-level false-positive detection is the consequence of the auto-classifier incorrectly assigning the focal species label to at least one recording. To proceed with standard occupancy modeling, one option is to effectively override or remove false-positive detections prior to statistical analyses (Barré et al., 2019). Currently, for North American bats, false-positive detections are removed either by using the “MLE-metric” from the computer software at a nightly level (visit-level) (e.g., Nocera et al., 2019b; Rojas et al., 2018), or by way of human vetters that are considered “experts” (Rojas et al., 2018; Banner et al., 2018). The human vetters follow a regional “key” or established rule-set to confirm at least 1-5 recordings are assigned the correct species label (e.g., Reichert et al., 2018). Occupancy modeling provides explanatory inferences about species probability of occurrence and how those may relate to environmental gradients, such as elevation, percent forest, and terrain ruggedness (e.g., Wright et al., 2021).

However, summarizing ARU data to a binary response may lack the resolution required for detecting more subtle population impacts from disease or disturbances (e.g., fire), creating temporally non-analog conditions that are spatially heterogeneous across a landscape (e.g., Law et al., 2018; Nocera et al., 2019a; Hyzy et al., 2020).

Recently, a Bayesian hierarchical model was developed for ARU data that provides estimates of species-specific activity at a local deployment location and estimates of site-level species occurrence associations with uncertainty (Wright et al., 2020; Stratton et al., in press 2022). The multi-species count detection model by Wright et al. (2020) utilizes the count of detections per species and avoids reducing the data to simply detection/non-detection for a nightly recording interval. The multi-species count detection model provides a potential bridge between two important objectives—broad-scale status and trends *and* local-scale decision making. In fact, we show the mathematical connection of the MLE-metric calculation (Britzke et al., 2002) and a posterior probability from the multi-species count detection model that allows for a direct comparison of these two methods for local-scale decision making (full details in Appendix S1). The key distinction between these approaches is that the count detection model allows for imperfect detection and inclusion of habitat and environmental conditions that inform and have the potential to improve determination of local presence for rare species.

We conduct a simulation-based assessment to compare the current USFWS regulatory standard of using the MLE-metric for determining species presence at a site to the multi-species count detection model and two alternative implementations of single-species occupancy models. We focus on the relative performance of the analytical approaches in terms of making a correct decision regarding species presence or absence at a specific location. A

concern when relying on statistical estimation for making a decision is whether the fitted models recover the data-generating values with reasonable accuracy and precision; if not, the ability to make a correct decision may be affected. Therefore, we include an investigation into parameter estimation for the Bayesian modeling approaches to verify whether they are statistically sound alternatives for application in local decision-making. Our work is motivated by our belief there is an unrealized opportunity in North American bat conservation to inform both site-level decisions about rare species presence **and** estimate species occurrence and relative activity related to environmental conditions using a unified statistical design and modeling framework.

2. METHODS

In Section 2.1, we outline the two primary sampling designs used in North America for collecting summertime bat acoustic data. The guidance differs because the intended spatial scope of inference varied: a specific project area versus rangewide extents (Fig. 1 denoted as Goal of Analysis). In Section 2.2, we provide details on the current USFWS standard for determining local species presence based on the MLE-metric (denoted as grey pathway in Fig. 1). Then in Section 2.3, we outline three Bayesian model-based approaches for bat ARU data that we include because of their potential for estimating the probability a species occurs locally and throughout its range (denoted as black pathways in Fig. 1). All the model-based options account for imperfect detection, but they differ in how species misclassifications are accounted for when estimating probability of occurrence (see Fig. 1). Finally in Section 2.4, we describe our simulation study design for comparing the four analytical approaches in

terms of their ability to assign the correct state of species (present or absent) at a site. The various factors we investigate were chosen because of their potential to influence estimating model parameters or assigning the (in)correct state of species presence or absence at a site (Fig. 1 denoted by dashed ovals with call-out boxes).

2.1 Summer Acoustic Bat Survey Guidelines

In North America, State and Federal regulations and policies require surveys to determine local presence of at least five bat species or sub-species (*Corynorhinus townsendii ingens*, *C. t. virginianus*, *Myotis grisescens*, *M. septentrionalis*, and *M. sodalis*) during the summer maternity season when most vulnerable to disturbance from management actions, e.g., forest harvesting, prescribed burning, surface mining, road construction and development (Silvis et al., 2016). Summer survey guidelines for determining sensitive species probable presence or absence to inform localized decisions are available from the U.S. Fish and Wildlife Service (USFWS, 2020). Currently, USFWS guidelines are geared towards two particular at-risk species that hibernate during the winter and day-roost in forests in the summer, *M. septentrionalis* and *M. sodalis*. Data are collected using stationary ARUs, but the spatial unit of a site is approximately 0.5 km² in area or along a 1 km linear unit. Deployment periods vary by target species and location, but generally are 8-9 precipitation-free nights for an areal project and 2 nights per km for linear. Typically, these data are then analyzed using the MLE-metric at the visit-level to make a decision regarding local species presence (Fig 1, grey pathway) before obtaining project clearance. We describe the MLE-metric calculation in Section 2.2.

Alternatively, an omnibus North American Bat Monitoring Program (NABat) designed to provide summertime status and trend estimation rangewide for multiple species simultaneously uses a common probabilistic master sample survey design for site selection (Larsen et al., 2008). The master sample is based on a grid-based sample frame with 10-km x 10-km cells. Within each selected NABat grid-cell, between 2 to 4 locations are selected for deploying ARUs for one to four nights (Loeb et al., 2015). The number of nights and locations within a grid cell can be informed by estimated species detection probabilities and field logistics (e.g., Rodriguez et al., 2019). NABat data have been analyzed at state-level (Neece et al., 2019), regional (Rodhouse et al., 2019), and rangewide (Udell et al., 2022) extents using various Bayesian hierarchical models. The inferential goals are typically to predict occurrence probabilities at all surveyed and non-surveyed sites and also estimate species-environment relationships with uncertainty. We consider a subset of the available options for modeling bat acoustic data (described in Section 2.3) for our simulation investigation because they also afford a pathway to site-level decisions about presence (Fig. 1, black pathways).

2.2 “MLE”-metric calculation

Britzke et al. (2002) introduced a MLE-metric for determining site-level species probable presence based on the number of recordings labeled to a focal species by the auto-classifier (hereafter, “autoID”) during a given recording interval (e.g., one night or aggregated over many nightly surveys). The approach explicitly acknowledges that autoIDs are subject to false positives, and a false positive for one species is a direct result of the presence and

incorrect identification of a recorded echolocation sequence from another species. Britzke et al. (2002) constructed a likelihood ratio (LR) test for the null hypothesis of species absence as an approach to decide species presence after accounting for software inaccuracies.

Following Britzke et al. (2002),

$\phi_{k'k}$ = probability that a recording from species k is identified to species k' . For a community of K species, a matrix with dimensions $K \times K$ is defined, where each row represents the software autoID result and each column is the true species identity,

θ_k^B = relative frequency of species k in sampled community, where we use the superscript B on the parameter to denote that it is from the Britzke et al. (2002) method,

N = total number of high-frequency sequences recorded and identified, and

n_k = the number of recordings identified to species k (autoIDs).

The classification probabilities ($\phi_{k'k}$) are assumed known and are based on a large set of voucher search-phase calls. From Britzke et al. (2002) “voucher calls were recorded from free flying bats in open areas with chemical light sticks on their backs such that the species emitting the echolocation call was known.” The relative frequency of a species in the sampled community (θ_k^B) are unknown parameters. The MLE-metric assumes a multinomial data likelihood for the total number of identifiable recordings and the number of autoIDs per species (n_1, \dots, n_K) with probabilities $\sum_{k=1}^K \phi_{k'k} \theta_k^B$;

$$L(\mathbf{n}, \theta) \propto \prod_i \left(\sum_{k=1}^K \phi_{k'k} \theta_k^B \right)^{n_k}. \quad (1)$$

The MLE-metric is the p-value from the likelihood ratio test with null hypothesis that

species k is absent from a site, $\theta_k^B = 0$ in Eqn. 1, versus the alternative hypothesis that the species is present, $\theta_k^B > 0$ in Eqn. 1. A p-value provides evidence against the null hypothesis — the probability of observing a test statistic at least as extreme (or more extreme) as what was observed, assuming the null hypothesis (species absence) and assumptions listed in Appendix S1 are reasonable.

Although the USFWS *M. sodalis* protocol determines species probable presence or absence as the MLE-metric applied at a visit-level (USFWS, 2020), we aggregated counts of autoIDs over all nightly recording sessions at a deployment location to both increase the number of recordings (sample size) and for an equivalent comparison to the model-based site-level decisions (denoted, “*MLESite*”; Appendix S1 and see Sec. 2.4.2). Current USFWS guidance suggests using a conservative 0.05 as the cut-off in the decision-rule for determining localized presence (if MLE-metric p-value < 0.05 , claim species present; USFWS, 2020). Hereafter, we denote the *MLESite* threshold value as α because it represents the significance level chosen for the hypothesis test (see Fig. 1).

2.3 Bayesian Models for Bat Acoustic Data

We explored three Bayesian modeling options for bat acoustic surveys (denoted by rounded rectangles in Fig. 1): 1) a standard single-species occupancy model that accounts for imperfect detection (denoted, “*Naive*”), 2) a single-species occupancy model with false-positive detections removed prior to analysis (denoted, “*Remove*”), and 3) a multi-species count detection model. The *Naive* and *Remove* approaches estimate probability of presence and detection for each species separately. Following notation from MacKenzie et al. (2002),

194 let $i = \{1, \dots, n\}$ index sites, $j = \{1, \dots, J\}$ visits. Then, the partially observed state of a
 195 site, Z_i is modeled as a Bernoulli random variable with probability ψ_i ($Z_i = 1$ if the focal
 196 species occurs at site i ; 0 otherwise). Presence (the Z -state) is imperfectly observed. The
 197 probability the focal species is detected at a site in which it occurs is denoted by p . Thus,
 198 $Y_{ij}|z_i = 1 \sim \text{Bernoulli}(p_{ij} * z_i)$, where y_{ij} is 1 if the focal species is detected at site i dur-
 199 ing visit j and 0 otherwise. Because $y_{ij} = 1 \implies z_i = 1$, the single-species occupancy
 200 model assumes no false-positive detections are made at the visit-level. Both the detection-
 201 level and occupancy-level of the model can be extended to include covariates that explain
 202 heterogeneity in p and ψ among visits or among sites (MacKenzie et al., 2002).

203 The difference between the *Remove* and *Naive* approaches lies in the construction of
 204 the detection history matrices (MacKenzie et al., 2002). The detection history matrix is
 205 defined as the $n \times J$ matrix of the observed y_{ij} values for all (i, j) . The *Remove* approach
 206 pre-processes the autoID counts for each species prior to fitting a single-species occupancy
 207 model. The pre-process step is achieved by implementing the MLE-metric with $\alpha = 0.05$
 208 based on the total number of identified recordings per species during a visit in an attempt to
 209 “*Remove*” false-positive errors prior to building the detection history matrix. The detection
 210 history matrix had a ‘1’ entered for a visit, if the MLE-metric is < 0.05 , otherwise a ‘0’ was
 211 entered. Alternatively, the *Naive* approach assumes that if at least one autoID is assigned to
 212 the focal species during a visit, then that species was present during the visit (i.e., a ‘1’ entry,
 213 otherwise a ‘0’ entry). The *Naive* approach incorrectly assumes the software auto-classifier
 214 always assigns the correct species label to a recording.

215 The multi-species count detection model, a Bayesian hierarchical model, was developed
 216 for ARU data to account for the same properties noted by Britzke et al. (2002) that the

217 rate of false positives for a species was related to both the autoclassification error rates and
 218 the presence and relative activity of other co-occurring bat species. In our application of
 219 the multi-species count detection model, we assume a two-species case (denoted, “*2sppCt*”).
 220 We assume the same information required for the MLE-metric is available for use in the
 221 *2SppCt* model. Specifically, we assume that the accuracy rates from the classifiers are known
 222 and correct. Appendix S1 provides the mathematical connection between the *MLESite*
 223 approach and the posterior probability from the *2SppCt*, which allows for a comparison of
 224 their (in)correct decisions about local presence within a simulation environment.

225 We briefly describe the multi-species count detection model, but see Wright et al. (2020)
 226 and Stratton et al. (in press 2022) for additional guidance on data requirements and Bayesian
 227 prior specification when the classification probabilities are assumed unknown and estimated
 228 jointly. Following notation from Wright et al. (2020), let $i = \{1, \dots, n\}$ index sites, $j =$
 229 $\{1, \dots, J\}$ visits to each site within a season, and $k = \{1, \dots, K\}$ index possible species
 230 available for recording during a study. For site i , species k was present (1) or not (0) with
 231 probability ψ_{ik} and we model this latent presence state as

$$Z_{ik} \sim \text{Bernoulli}(\psi_{ik}), \quad (2)$$

232 then an appropriate link function (e.g., logit, probit) can be used to model site-level infor-
 233 mation to inform the probability a species occurred at a site ($Z_i = 1$) such as elevation or
 234 percentage forest cover. Given species k occurs at site i , the true number of echolocation

235 sequences (recordings) for a species is modeled as,

$$[Y_{ijk} \mid Z_{ik} = 1] \sim \text{Poisson}(\lambda_{ijk}), \quad (3)$$

236 where λ_{ijk} is the expected (mean) number of recordings per night and $\log(\lambda_{ijk})$ can be
 237 used to include covariates associated with relative bat activity during a site-night visit, such
 238 as nightly minimum temperature or average wind speed. In the count detection model, the
 239 probability of a false negative (zero recorded and/or identifiable calls even though species
 240 present at site) is estimable as, $Pr(Y_{ijk} = 0 \mid Z_{ik} = 1) = \exp(-\lambda_{ijk})$. The observed re-
 241 sponse data are the sum of autoIDs for each species (n_k) and are modeled as Poisson counts,
 242 $[n_k \mid \mathbf{Z}, \boldsymbol{\lambda}, \boldsymbol{\phi}] = \text{Poisson}(\sum_{k'=1}^K Z_{ik'} \lambda_{ijk'} \phi_{kk'})$.

243 2.4 Simulation Study Methods

244 For our study, we assumed only two species because our exploration utilized computer
 245 simulations and increasing to more species increased the computational demand. Our previ-
 246 ous work verified that the multi-species count detection model returns unbiased parameter
 247 estimates when extended to bat assemblages composed of ten species (Stratton et al., in press
 248 2022). Similarly, other work has established the overestimation of occupancy probabilities
 249 when the *Naive* model is applied to datasets contaminated by false-positive detections from
 250 an omnibus source (Royle and Link, 2006), but we include it for completeness. The *Remove*
 251 approach has not been evaluated previously using a statistical simulation study, only with
 252 empirical observations of *M. septentrionalis* (Rojas et al., 2018).

253 Our first objective was to compare the three model-based approaches for estimating both

the probability a species is detected (p) and the probability a species occurs (ψ). Our second objective was to compare the *MLESite* and suitable model-based approaches in terms of their application for discerning site-level species presence or absence correctly within the same simulation framework. Because we know that automatic classifiers for bat acoustic data are not 100% accurate, we generated datasets of autoIDs contaminated by recording or observation-level identification errors for two species (data were consistent with the *2SppCt* model). The chosen parameter settings reflected our current understanding of two species (*M. lucifugus* and *M. sodalis*) that produce similar search-phase echolocation calls and tend to co-occur, and display considerable overlap in habitat association for many areas (Ford et al., 2005, Table 1). One species was assumed rare (e.g., *M. sodalis*) within a study area ($\psi_1 = 0.25$) and locally less active (relative activity rates averaging around 1 identifiable recording per night). The second species was assumed more widespread (e.g., *M. lucifugus*; $\psi_2 = 0.75$) and more active (relative activity rates with on average 10 identifiable recordings per night).

We considered five different simulation scenarios. Three scenarios reflected different (miss)-classification probabilities with varying contributions of false positive recordings from the common species (ϕ , S1, S2, S3: Table 1). Scenario 1 assumed that the auto-classifier was 90% accurate for rare species recordings and only 65% accurate for a common species; consequently, the contribution of false positives from the common species were considerable at 35%. Scenario 2 assumed that the auto-classifier was only 65% accurate for rare species recordings, but the false positive contributions from a common species were lower with only 10% of recordings of the common species misidentified. Scenario 3 assumed the auto-classifier accuracy rate for both the rare and common species was low, respectively 65% and 60%. The

remaining two scenarios (S4 and S5) provided direct comparisons to Scenario 1, the classifier accuracy rates were assumed the same (ϕ equal for S1, S4, and S5 in Table 1). Scenario 4 assumed a higher occurrence probability for the rare species ($\psi_1 = 0.50$, Table 1). Scenario 5 assumed higher relative activity of the rare species, thereby increasing the expected number of recordings classified during a nightly interval.

All scenarios were assumed to have 55 sites with 8 or 16 recording nights per site for consistency with current guidelines from NABat and USFWS for the endangered *M. sodalis* and the *M. lucifugus*, a candidate for possible listing. We simulated 50 datasets under each scenario for both 8 and 16 visits, and applied all four analytical approaches to each dataset. Again, for the MLE-metric calculations and the *2SppCt*, we assumed that the auto-classifier (in)accuracy rates were known without error. The three model-based approaches were fit using NIMBLE (de Valpine et al., 2017, 2021). The different scenarios required different iterations, warm-up, and thinning of the MCMC to exhibit good mixing of the chains and evidence of convergence. Chains were visually assessed using traceplots, summarized with the Gelman-Rubin diagnostic (*Rhat*) and the number of effectively independent samples (n_{eff}) (*Rhat*, n_{eff} computed using **rstan**, Stan Development Team, 2020). Details for each scenario are provided in Appendix S2. We assumed diffuse priors of $[\psi_k] = Uniform(0, 1)$; $[p_k] = Uniform(0, 1)$; and a weakly informative prior of $[\lambda_k] = Gamma(shape = 2, rate = 0.25)$, which is right skewed with the majority of the mass between 0 and 20 and a mean of 8 (median 6.7) and represents a reasonable range of relative activity for the species we consider.

2.4.1 Parameter Estimation

We compared estimated probability for rare species presence (ψ_1) and common species presence (ψ_2) among the *2SppCt*, *Remove*, and *Naive* approaches, as all three models provided Bayesian posterior distributions for these parameters. For visualization, we reported the average of the 2.5%, 50%, and 97.5% percentiles from the parameter posterior distributions over the 50 iterations of simulated data for all three models. We assessed estimation error (bias) and relative uncertainty conveyed by the 95% posterior intervals for all three models. Coverage was computed by calculating the proportion of simulated datasets that resulted in 95% posterior intervals including the true data-generating values (*Note*: slight deviations from nominal coverage are expected because of the small number (50) of simulated datasets for each scenario). The *Remove* and *Naive* occupancy models provide an estimate of detectability (p_1 and p_2 for both species), whereas the *2SppCt* model provides an estimate of species-specific relative activity (λ_1 and λ_2).

2.4.2 Assessing site-level decisions regarding species presence (absence)

For each simulated dataset, we determined a site-level decision regarding a conclusion of species absent or present with all four approaches (*MLESite*, *2SppCt*, *Remove*, *Naive*) for comparison to the true known state of species absent ($Z = 0$) versus present ($Z = 1$). We summarized the number of sites with a specific decision conditional on the known true state of species presence or absence for each dataset comprised of 55 sites.

The *MLESite* approach used the total number of autoIDs classified to each species aggregated over the nightly surveys (8 or 16 visits) to arrive at a site-level decision of species

presence or absence. The *MLESite* decision-rule is based on a threshold that is related to a maximum tolerance for making a Type I error (claiming species was present when actually absent). We explored different thresholds (α) for the decision-rule embedded in the *MLESite* approach for concluding species presence (MLE-metric $< \alpha$) or absence (MLE-metric $\geq \alpha$). We considered $\alpha = 0.05, 0.1, 0.15$, and 0.2 , which represents a reasonable range of tolerances for making Type-I errors (erroneously concluding species presence). These results are presented in Appendix S2.

The model-based approaches (*2SppCt*, *Remove*, *Naive*) arrived at a site-level decision by using the Bayesian posterior probability a species occurs at a site. The model-based approaches used the posterior mean of the Z-state for each site ($Pr(Z_i = 1|y)$) compared to a threshold (z_{cutoff}) to make a site-level decision about species presence or absence. The decision-rule was defined as, $Pr(Z_{ik} = 1|y) > z_{cutoff} \implies$ species k was claimed present at site i , otherwise species k was claimed absent at site i . We considered z_{cutoff} values of $0.05, 0.25, 0.5, 0.75$, and 0.95 to represent a range of decision-rules. Smaller z_{cutoff} values increased the number of sites with a decision of species present. Conversely, larger values of z_{cutoff} increased the number of sites with the state of species absent assigned. The threshold investigation results are presented in Appendix S2.

3. Results

Generally, we found that changing the classifier accuracy from 0.90 to 0.65 and recording-level false-positive rate from 0.35 to 0.10 for the rare species did not impact *2SppCt* parameter estimation or the site-level decisions of *MLESite* and *2SppCt* because we assumed those val-

ues were known and correct (results in Appendix S2). The scenario where both species had poor auto-classifiers (true classification rates of 0.65 for rare species and 0.60 for common species in Scenario 3, Table 1) led to poor estimation of the relative activity using *2SppCt* across the 50 simulated datasets (Appendix S2). The unstable results suggest a minimum accuracy for species classification should be established, which was consistent with the rigorous testing required by the USFWS to approve software for rare species surveys (USFWS and USGS, 2019).

We include all results for the common, easily detected species in Appendix S2. The results showed little variability among scenarios and near perfect correct site-level decisions among the 5 data-generating scenarios we investigated. This is not very surprising because site-level decisions about species presence are robust when a species has very high probability of being detected acoustically, when present. The contributions of recording-level false-positives from the rare species were too infrequent relative to the number of correct identifications of the common species to bias the decision outcome. The utility of including the “common” species, statistically, is as the source of recording-level false-positives for the rare species.

In the following sections, we explore parameter estimation for the model-based approaches (Section 3.1) and then compare *MLESite* approach to suitable model-based approaches for determining whether the rare species was present at a site (Section 3.2). We assume the classifier was 90% accurate identifying the rare species with a substantial source of false positives contributed by the common species recordings (35%). We focus on comparing three scenarios to investigate whether increasing the occurrence probability for the rare species from $\psi_1 = 0.25$ to $\psi_1 = 0.5$ (denoted, $[\psi = 0.25, \lambda = 0.3]$ versus $[\psi = 0.5, \lambda = 0.3]$) or increasing the detection probability for the rare species from $p_1 = 0.25$ to $p_1 = 0.5$ (denoted,

[$\psi = 0.25, \lambda = 0.3$] versus [$\psi = 0.25, \lambda = 0.7$]) influenced model parameter estimates or site-level decisions.

3.1 *Naive, Remove, and 2SppCt* model estimation for rare species occurrence and detection probabilities

The count detection model is the most realistic data-generating model for bat acoustic data currently in the literature, and given that simulated datasets were generated under this model, it was not surprising that the *2SppCt* version consistently returned unbiased ψ estimates for the rare species (Fig. 2 top row with additional results in Appendix S2: Figures S1-S2). The posterior estimates for relative activity or the average number of recordings per species were unbiased and 95% coverage was achieved for most scenarios, as expected (Fig. 2 bottom row and Appendix S2: Figure S1-S2). There was a slight reduction in uncertainty (narrower posterior intervals) for the rare species with 16 rather than 8 visits (Appendix S2 Figures S1-S2).

The *Remove* approach overestimated probability of presence (Fig. 2 top row) and underestimated the probability of detection regardless of the assumed average occurrence or relative activity (Fig. 2 middle row). The small detection probabilities explains the larger uncertainty for occurrence probabilities using *Remove* (Fig. 2). The posterior intervals based on the *Remove* approach were more noticeably reduced by doubling the number of visits at a site from 8 to 16 (Appendix S2 Figures S1-S2).

The *Naive* approach consistently overestimated rare species occurrence probability (Fig. 2 top row) and probability of detection (Fig. 2 middle row) for all scenarios because of

false-positive detections contributed from a common species. These parameter estimates translated to an “always present” decision for every site. The severe bias in parameter estimation using the *Naive* occupancy model suggested its application for assessing rare bat species presence requires caution and justification that no false-positives occur.

3.2 *MLESite* versus *Remove* or *2SppCt* model performance for determining local species presence

We present our investigation into whether the threshold value specified in the decision-rule influenced the median proportion of sites assigned the correct state of rare species present or absent in Appendix S2. We used cut-off values that balanced the correct versus in-correct decision rates regarding species presence or absence at a site in the comparisons we highlight (Fig. 3). The *MLESite* decision-rule for species presence was p-value < 0.10 and p-value > 0.10 for species absence at a site. The posterior probabilities from *2SppCt* and *Remove* were used with a threshold specific to the desired decision of species presence or absence: decision-rule of $Pr(Z = 1|y) > 0.75$ for species absence (protecting against erroneous determination of species presence) and a decision-rule of $Pr(Z = 1|y) > 0.25$ for species presence (protecting against erroneous determination of species absence) at a site.

Regardless of the assumed occurrence and relative activity for the rare species in our simulations, the *2SppCt* model outperformed the *MLESite* approach for correctly determining a species was present at a site (Fig. 3 top left panel). The *2SppCt* model median correct decision-rate for presence increased from 0.78 to 1 as the occurrence probability increased from 0.25 to 0.50, and the variability in correct decision-rates decreased. The *MLESite* me-

dian correct decision-rate for presence was improved when the assumed average number of recordings increased from 5 to 11 over the entire survey duration (16 nights of recording). However, the median proportion of correct decisions for presence was always less than 0.5 for the *MLESite* (Fig. 3. top left panel). The *Remove* approach always returned a decision of species presence, regardless of the data-generating values (Fig 3. top left panel).

Another way to explore the performance for informing local decisions is to consider the median proportion of sites with incorrect decisions of species absent when truly present (Fig. 3 bottom left panel, notice bottom panel is essentially a mirror image of the corresponding top panel). The *2SppCt* model consistently had the lowest median incorrect decision-rates compared to *MLESite* and *Remove* approach regardless of the assumed relative activity or occurrence probabilities (Fig. 3 bottom left panel). Again, the performance for the *2SppCt* model improved when the assumed occurrence probability for a rare species was 0.50 and the *MLESite* incorrect decision-rate was reduced by increasing the expected number of nightly recordings. The *Remove* approach infrequently resulted in a decision of species absence (very few points graphed in bottom left panel Fig. 3).

For all the approaches we explored, the correct decision-rates for species absence were less variable with no clear differences related to assumed relative activity or occurrence probabilities (Fig. 3 right panels). The *2SppCt* model was equivalent to or better than the *MLESite* approach for determining a rare species was absent at a site with both having median correct decisions $> 90\%$ of the time (Fig 3., bottom right panel). The *Remove* had the lowest correct decision-rates for species absence (Fig 3. lower right panel).

Both the *2SppCt* and *MLESite* had consistently the lowest median incorrect decision-rate for claiming a rare species was present when truly absent (Fig. 3, top right panel).

Note the *MLESite* had a median incorrect decision-rate of approximately 10%. As expected, the number should be close to the cut-off value or α used to make a decision because the *MLESite* was based on a frequentist hypothesis test. The *Remove* approach had consistently the highest median incorrect decision-rate (approx. 25%) regarding local absence because most sites were assigned the incorrect state of rare species presence (Fig. 3, top right) due to the bias in the occurrence and detection parameter estimates (Fig. 2, top two rows).

4. Discussion

Our simulation investigation indicated that the multi-species count detection model (in the two species case) was equivalent to or better than the current USFWS standard for determining local species presence, which opens up possibilities for integrating summertime acoustic datasets within a common statistical inference framework. Below, we outline the key findings for the four approaches (*MLESite*, *2SppCt*, *Remove* and *Naive*) compared under realistic data-generating scenarios for bat ARU data. Then based on our simulation findings, we describe future field investigations that, in coordination with statistical method development, could provide huge dividends for bat conservation. We conclude with a discussion on statistical considerations prior to combining datasets from the different summertime acoustic surveys for rare bats and provide suggestions for sound data integration and Bayesian inferences with uncertainty in future.

We found that applying the MLE-metric at a site-level (*MLESite*) had median correct decisions less than 60% for species presence with most scenarios ranging between 20% and 30%, but greater than 85% for species absence (ranging between 80% and 90%). The *MLE-*

Site had reasonable correct decision-rates for absence consistent with its intended purpose of null hypothesis testing for species absence (Section 2.2). However, to increase the correct decision-rate for species presence requires increasing the threshold-value (α) and/or increasing the total number of identified recordings (N) and the expected number of recordings identified to each species ($E(n_k)$ in Eqn. 1; de Jong et al., 2019). Practically, a larger sample size could result from aggregating recordings over more visits to a site, orienting ultrasonic microphones to reduce interference that hampers recording quality (Loeb et al., 2015, chapter 4), and/or strategic placement of detectors in preferred habitats of the focal species within the project area (e.g., for *M. sodalis* and *M. septentrionalis* upland and riparian forested conditions; Menzel et al., 2001; Silvis et al., 2016). However, data collected using such a preferential sampling design are no longer representative of a larger collection of sites (statistical population). Although the *MLESite* accounts for observation-level misclassifications, the information output is only a binary decision of species presence or absence and potential sources of meaningful ecological variation and imperfect detection are ignored (Fig. 1, denoted by grey pathway).

We propose an alternative to *MLESite* is to consider the multi-species count detection model. We found, in the two species case, that *2SppCt* had equivalent or greater correct claims of rare species presence (median > 75%) or absence (median > 80%) at a site. The count detection model was developed with the same underlying assumptions as *MLESite*, but can be modified and tuned for different applications. For example, variables that characterize differences in recording conditions (e.g., nightly temperature Gorman et al., 2021) or deployment locations (e.g., near water features, flyway, interior forest, etc; Ford et al., 2005) can be exploited to adjust inferences appropriately for nightly recording sessions that had no

or very few bat calls detected. Depending on the spatio-temporal sampling design underlying an application, additional parameters can be included that model spatial correlation in probability of occurrence (Eqn. 2) and/or temporal correlation in relative activity (Eqn. 3) using Gaussian processes (Gelfand and Schliep, 2016; Wright et al., 2021). Importantly, the count detection model allows for deeper insights into species-environment relationships in **both** relative bat activity and presence. These estimated associations, when underpinned by a probabilistic survey design, could be used to improve prediction of rare species occurrence at a site thereby increasing the utility of bat ARU surveys to fulfill the goals of ecological research and regulatory compliance information needs.

Our findings are consistent with previous work that urges caution in the application of the *Naive* single-species occupancy model for ARU data subject to observation-level false positives introduced during the automated classification process (Chambert, Campbell Grant, Miller, Nichols, Mulder and Brand, 2018). The severe overestimation of occurrence and detection probabilities translated into essentially a conclusion of the rare species always being present because the common species was always available to contribute false-positive detections. We suggest the *Naive* approach should only be considered when the autoclassifier is 100% accurate for all species that emit potentially overlapping echolocation sequences. Even a small source of false-positive detections ($< 5\%$) can bias occupancy model parameter estimates and, consequently, lead to an incorrect conclusion that a species is present within an area (Miller et al., 2013; Chambert et al., 2015)

However, applying the MLE-metric at the visit-level to “remove” false-positives prior to model fitting (i.e., the *Remove* approach) resulted in many sites where the species truly occurred with zero detections. Our findings that occurrence probabilities were overestimated

and detection probabilities severely underestimated were consistent with field-based comparisons of our *Remove* occupancy model to site confirmation false-positive models for *M. septentrionalis* (Rojas et al., 2018). As a simple demonstration of the ramifications parameter estimation bias can have on site-level decisions; a site with zero detections and the data-generating values we specified in Scenario 1 would have 0.003 posterior probability that a species occurred (Table 1; MacKenzie et al., 2002). However, using the biased average posterior estimates based on the *Remove* model with 16 nights, the probability would be 0.40. An always present decision is perhaps satisfactory from a conservative conservation stand-point to ensure the minimization of potential “take,” but also may incorrectly attribute habitat associations or correlates to a rare species that might divert or misallocate habitat management and protection efforts. Based on our simulations, relying on the MLE-metric at a visit-level as a means to remove false-positives detections was not sufficient for modeling bat acoustic datasets; however, application in real-world conditions with potentially a greater number of identified recordings and a larger bat species pool could present a different perspective.

Our findings suggest the need for follow-up field studies that compare the *MLESite*, *Remove* approach, and the multi-species count detection model for determining species presence or absence using empirical data with a larger species assemblage. In our simulations, we assumed the true classifier accuracy rates were known and constant for all visits and sites. If these assumptions are not met the parameter estimates for λ and ψ would be biased and decisions based on the *MLESite* and count detection model potentially compromised. Curating local to regional calibration or confirmation datasets that tune the classification probabilities to real-world recording conditions could substantially improve rare species deci-

sions and provide significant cost reductions for studies that employ human experts to verify the assigned species to a recording (autoIDs; see Stratton et al., in press 2022). Ideally, future work would consider co-locating ARU-based surveys with other field methods (known roosts or mist-netting) that provide near definitive evidence that a rare species was actually present (e.g., Miller et al., 2015; Rojas et al., 2018). Pairing such short-term field investigations with statistical method development increases the potential for both lines of inquiry to improve conservation decision-making for rare bat populations.

Our simulations demonstrate that the multi-species count detection model holds promise as a cohesive statistical inference framework for leveraging both local, targeted data collection with rangewide probabilistic sampling to inform regional assessments and could reduce uncertainty regarding local species presence. However, prior to integrating empirical datasets that are collected under different objectives and survey protocols, the potential misalignment in how sites are selected (spatial design), the grain size of a site (spatial resolution), and the guidance on how field data are collected (response design) should be considered thoroughly. For summertime bat acoustic surveys in North America, the main difference between the NABat and USFWS protocols is the definition of a “site” (Sect. 2.1) because of the different intended objectives. The omnibus NABat program defines a “site” as a 10-km \times 10-km grid cell based on average assumed dispersal distances (Loeb et al., 2015). For regulatory project clearance prior to conducting activities within suitable summertime *M. sodalis* and *M. septentrionalis* habitat, a site is defined as a 0.5-km² area (USFWS, 2020). The difference in analytical unit should not prevent integrating datasets gathered under the different sampling designs. In fact, a straightforward extension of the count detection model to a multi-scale parameterization that estimates an availability parameter and specifies an inter-

mediate latent-level for rare species presence/absence at a specific detector location (e.g., Nichols et al., 2008) provides a reasonable statistical inference framework for a combined dataset. The same model would allow for estimating grid-cell level species occurrence and information from other survey sites and could be leveraged to improve the decisions made at specific locations within a grid cell.

Minimal alterations to the summertime acoustic survey designs are needed to achieve robust data integration of the regulatory survey data within the broader NABat framework. The NABat guidance already suggested having both spatial and temporal replication within a grid cell (Loeb et al., 2015). Therefore, the only potential modification would be to increase the number of nights to be consistent with USFWS guidance for **some** detector locations. Also, local site selection for which habitats are optimal for detecting sensitive species need not be changed, the information (metadata) characterizing the habitat surrounding the detector location can be included in an integrated model as covariates for relative activity and local availability, thereby adjusting for the preferential detector placement within a grid cell. For example, the putative day-roosting and foraging habitat of *M. septentrionalis* and *M. sodalis* in eastern North America is upland and riparian forested conditions (Gorman et al., 2022). The NABat acoustic survey design suggests selecting different habitat features within a grid cell to maximize the potential for detecting the full suite of species. The practical hurdle is properly measuring and archiving the local habitat in a consistent manner among survey efforts. The NABat community of practice could provide the mechanism for developing consistent and compatible data fields about ARU summertime surveys and the online data submission portal facilitates the upload and download of both the design and response data for species assessments (Reichert et al., 2021).

In addition to the follow-up field investigations that we outlined and the clear potential for data integration of summertime acoustic survey data within the multi-species count detection modeling framework; we suggest an opportunity exists to exploit a full Bayesian statistical decision analysis (BDA) for rare species decision-making. A BDA combines the posterior probability that a species was present with a loss or risk function to determine the potential choice (claim species present or claim species absent) that has the lowest expected risk (Williams and Hooten, 2016). A loss function could be developed *with* decision- and policy-makers input regarding whether an erroneous claim of species absence or species presence was more detrimental across a suite of stewardship actions and outcomes (Wade, 2000). For example, the economic cost of halting or modifying a forest management action or an infrastructure project due to an incorrect decision of rare species absence versus the potential loss of proceeding and increasing the possible accidental “take” for an endangered species, if the species was truly present. The decision-rule for species presence would be based on the decision outcome with the lowest expected risk. However, for some situations there could be little distinction between the two outcomes expected risk values, but the full uncertainty in the data and a clear articulation of the losses provide added transparency in decision-making.

5. Acknowledgements

We thank Ted Weller, Cori Lausen, Dan Bachen, and Jason Rae for educating the statisticians about bat acoustic data workflows. Two anonymous reviewers’ comments substantially improved our manuscript. K. Irvine and K. Banner appreciate the Bayesian decision anal-

585 ysis virtual reading group participants’ observations during Spring 2021. K. Irvine and C.
586 Stratton’s participation was made possible by U.S. Geological Survey Ecosystems Mission
587 Area funding for Whitenose Disease and NABat Acoustic Research and Development (“Bats
588 and Stats”). Funding for K. Banner’s participation was supported by cooperative agreement
589 between U.S. Geological Survey and Montana State University, award No. G20AC00406.
590 Any use of trade, firm, or product names is for descriptive purposes only and does not imply
591 endorsement by the U.S. Government.

References

- Acevedo, M. A., Corrada-Bravo, C. J., Corrada-Bravo, H., Villanueva-Rivera, L. J. and Aide, T. M. (2009), ‘Automated classification of bird and amphibian calls using machine learning: A comparison of methods’, *Ecological Informatics* **4**(4), 206–214.
- Aodha, O. M., Gibb, R., Barlow, K. E., Browning, E., Firman, M., Freeman, R., Harder, B., Kinsey, L., Mead, G. R., Newson, S. E., Pandourski, I., Parsons, S., Russ, J., Szodoray-Paradi, A., Szodoray-Paradi, F., Tilova, E., Girolami, M., Brostow, G. and Jones, K. E. (2018), ‘Bat detective - deep learning tools for bat acoustic signal detection’, *PLOS Computational Biology* **14**(3), 1–19.
- Banner, K. M., Irvine, K. M., Rodhouse, T. J., Wright, W. J., Rodriguez, R. M. and Litt, A. R. (2018), ‘Improving geographically extensive acoustic survey designs for modeling species occurrence with imperfect detection and misidentification’, *Ecology and Evolution* **8**(12), 6144–6156.
- Barré, K., Le Viol, I., Julliard, R., Pauwels, J., Newson, S. E., Julien, J.-F., Claireau, F., Kerbiriou, C. and Bas, Y. (2019), ‘Accounting for automated identification errors in acoustic surveys’, *Methods in Ecology and Evolution* **10**(8), 1171–1188.
- Britzke, E., Murray, K., Heywood, J. and Robbins, L. (2002), *The Indiana bat: Biology and management of an endangered species*, Austin, Texas: Bat Conservation International, chapter Acoustic Identification, pp. 221–225.
- Chambert, T., Campbell Grant, E. H., Miller, D. A. W., Nichols, J. D., Mulder, K. P. and

Brand, A. B. (2018), ‘Two-species occupancy modeling accounting for species misidentification and nondetection’, *Methods in Ecology and Evolution* **9**(6), 1468–1477.

Chambert, T., Miller, D. A. W. and Nichols, J. D. (2015), ‘Modeling false-positive detections in species occurrence data under different study designs’, *Ecology* **96**(2), 332–339.

Chambert, T., Waddle, J. H., Miller, D. A. W., Walls, S. C. and Nichols, J. D. (2018), ‘A new framework for analysing automated acoustic species detection data: Occupancy estimation and optimization of recordings post-processing’, *Methods in Ecology and Evolution* **9**(3), 560–570.

Cheng, T. L., Reichard, J. D., Coleman, J. T. H., Weller, T. J., Thogmartin, W. E., Reichert, B. E., Bennett, A. B., Broders, H. G., Campbell, J., Etchison, K., Feller, D. J., Geboy, R., Hemberger, T., Herzog, C., Hicks, A. C., Houghton, S., Humber, J., Kath, J. A., King, R. A., Loeb, S. C., Massé, A., Morris, K. M., Niederriter, H., Nordquist, G., Perry, R. W., Reynolds, R. J., Sasse, D. B., Scafani, M. R., Stark, R. C., Stihler, C. W., Thomas, S. C., Turner, G. G., Webb, S., Westrich, B. J. and Frick, W. F. (2021), ‘The scope and severity of White-nose syndrome on hibernating bats in North America’, *Conservation Biology* **35**(5), 1586–1597.

Clement, M. J., Rodhouse, T. J., Ormsbee, P. C., Szewczak, J. M. and Nichols, J. D. (2014), ‘Accounting for false-positive acoustic detections of bats using occupancy models’, *Journal of Applied Ecology* **51**(5), 1460–1467.

de Jong, V. M. T., Eijkemans, M. J. C., van Calster, B., Timmerman, D., Moons, K. G. M., Steyerberg, W. and van Smeden, M. (2019), ‘Sample size considerations and predictive

performance of multinomial logistic prediction models’, *Statistics in Medicine* **38**(9), 1601–
1619.

de Valpine, P., Paciorek, C., Turek, D., Michaud, N., Anderson-Bergman, C., Obermeyer, F.,
Wehrhahn Cortes, C., Rodríguez, A., Temple Lang, D. and Paganin, S. (2021), ‘NIMBLE:
MCMC, particle filtering, and programmable hierarchical modeling’. R package version
0.11.0.

de Valpine, P., Turek, D., Paciorek, C., Anderson-Bergman, C., Temple Lang, D. and Bodik,
R. (2017), ‘Programming with models: writing statistical algorithms for general model
structures with NIMBLE’, *Journal of Computational and Graphical Statistics* **26**, 403–
413.

Erickson, R., Thogmartin, W. and Szymanski, J. (2014), ‘BatTool: an R package with gui
for assessing the effect of White-nose syndrome and other take events on *Myotis* spp. of
bats’, *Source code for biology and medicine* **9**.

Ford, W. M., Menzel, M., Rodrigue, J., Menzel, J. and Johnson, J. (2005), ‘Relating bat
species presence to simple habitat measures in a central Appalachian forest’, *Biological
Conservation* **126**(4), 528–539.

Frick, W. F., Kingston, T. and Flanders, J. (2020), ‘A review of the major threats and
challenges to global bat conservation’, *Annals of the New York Academy of Sciences*
1469(1), 5–25.

Gelfand, A. and Schliep, E. (2016), ‘Spatial statistics and gaussian processes: A beautiful
marriage’, *Spatial Statistics* **18**(A), 86–104.

- Gorman, K., Deeley, S., Barr, E., Freeze, S., Kalen, N., Muthersbaugh, M. and Ford, W. (2022), ‘Broad-scale geographic and temporal assessment of northern long-eared bat (*Myotis septentrionalis*) maternity colony landscape association’, *Endangered Species Research* **47**, 119–130.
- Gorman, K. M., Barr, E., Ries, L., Nocera, T. and Ford, W. (2021), ‘Bat activity patterns relative to temporal and weather effects in a temperate coastal environment’, *Global Ecology and Conservation* **30**, 1–13.
- Hoyt, J. R., Kilpatrick, A. M. and Langwig, K. E. (2021), ‘Ecology and impacts of White-nose syndrome on bats’, *Nature Review Microbiology* **19**, 196–210.
- Hyzy, B. A., Russell, R. E., Silvis, A., Ford, W. M., Riddle, J. and Russell, K. (2020), ‘Occupancy and detectability of northern long-eared bats in the lake states region’, *Wildlife Society Bulletin* **44**(4), 732–740.
- Larsen, D. P., Olsen, A. R. and Stevens, D. L. (2008), ‘Using a master sample to integrate stream monitoring programs’, *Journal of Agricultural, Biological, and Environmental Statistics* **13**(3), 243–254.
- Law, B. S., Doty, A., Chidel, M. and Brassil, T. (2018), ‘Bat activity before and after a severe wildfire in Pilliga forests: Resilience influenced by fire extent and landscape mobility?’, *Austral Ecology* **43**(6), 706–718.
- Loeb, S. C., Rodhouse, T. J., Ellison, L. E., Lausen, C. L., Reichard, J. D., Irvine, K. M., Ingersoll, T. E., Coleman, J. T., Thogmartin, W. E., Sauer, J. R., Francis, C. M., Bayless, M. L., Stanley, T. R. and Johnson, D. H. (2015), A plan for the North American Bat

Monitoring Program (NABat), General Technical Report SRS-208, U.S. Department of
Agriculture, Forest Service, Southern Research Station, Asheville, NC.

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A. and Langtimm,
C. A. (2002), ‘Estimating site occupancy rates when detection probabilities are less than
one’, *Ecology* **83**(8), 2248–2255.

Menzel, M., Menzel, J., Carter, T., Ford, W. and Edwards, J. (2001), Review of the forest
habitat relationships of the Indiana bat (*Myotis sodalis*), General Technical Report NE-
284, USDA Forest Service, Newtown Square, PA.

Miller, D. A. W., Bailey, L. L., Campbell Grant, E. H., McClintock, B. T., Weir, L. A. and
Simons, T. R. (2015), ‘Performance of species occurrence estimators when basic assump-
tions are not met: a test using field data where true occupancy status is known’, *Methods
in Ecology and Evolution* **6**(5), 557–565.

Miller, D. A. W., Nichols, J. D., Gude, J. A., Rich, L. N., Podruzny, K. M., Hines, J. E.
and Mitchell, M. S. (2013), ‘Determining occurrence dynamics when false positives occur:
Estimating the range dynamics of wolves from public survey data’, *PLOS ONE* **8**(6), 1–10.

Neece, B. D., Loeb, S. C. and Jachowski, D. S. (2019), ‘Implementing and assessing the
efficacy of the north american bat monitoring program’, *Journal of Fish and Wildlife
Management* **10**(2), 391–409.

Nichols, J. D., Bailey, L. L., JR., A. F. O., Talancy, N. W., Grant, E. H. C., Gilbert,
A. T., Annand, E. M., Husband, T. P. and Hines, J. E. (2008), ‘Multi-scale occupancy

estimation and modelling using multiple detection methods’, *Journal of Applied Ecology*
45(5), 1321–1329.

Nocera, T., Ford, W. M., Silvis, A. and Dobony, C. A. (2019*a*), ‘Patterns of acoustical
activity of bats prior to and 10 years after WNS on Fort Drum Army Installation, New
York’, *Global Ecology and Conservation* **18**, e00633.

Nocera, T., Ford, W., Silvis, A. and Dobony, C. (2019*b*), ‘Let’s agree to disagree: Comparing
auto-acoustic identification programs for northeastern bats’, *Journal of Fish and Wildlife*
Management **10**(2), 346–361.

Polasky, S., Carpenter, S. R., Folke, C. and Keeler, B. (2011), ‘Decision-making under great
uncertainty: environmental management in an era of global change’, *Trends in Ecology*
and Evolution **26**(8), 398–404.

Reichert, B., Bayless, M., Cheng, T., Coleman, J., Francis, C., Frick, W., Gotthold, B.,
Irvine, K., Lausen, C., Li, H., Loeb, S., Reichard, J., Rodhouse, T., Segers, J., Siemers,
J., Thogmartin, W. and Weller, T. (2021), ‘NABat: a top-down, bottom-up solution to
collaborative continental-scale monitoring’, *Ambio* **50**, 901–913.

Reichert, B., Lausen, C., Loeb, S., Weller, T., Allen, R., Britzke, E., Hohoff, T., Siemers, J.,
Burkholder, B., Herzog, C. and Verant, M. (2018), A guide to processing bat acoustic data
for the North American Bat Monitoring Program (NABat), Technical report, Open-File
Report 2018–1068, U.S. Geological Survey, Reston, Virginia, USA.

URL: <https://doi.org/10.3133/ofr20181068>

Rodhouse, T. J., Ormsbee, P. C., Irvine, K. M. and Vierlin, L. A. (2012), ‘Assessing the status

and trend of bat populations across broad geographic regions with dynamic distribution models', *Ecological Applications* **22**(4), 1098–1113.

Rodhouse, T. J., Rodriguez, R. M., Banner, K. M., Ormsbee, P. C. and Irvine, K. M. (2019), 'Evidence of region-wide bat population decline from long-term monitoring and Bayesian occupancy models with empirically informed priors', *Ecology and Evolution* **9**(19), 11078–11088.

Rodriguez, R. M., Rodhouse, T. J., Barnett, J., Irvine, K., Banner, K. M., Lonneker, J. and Ormsbee, P. C. (2019), North American Bat Monitoring Program regional protocol for surveying with stationary deployments for echolocation recording devices: Narrative version 1.0, Pacific Northwestern US, Natural Resources Report NPS/UCBN/NRR — 2019/1975, National Park Service, Fort Collins, CO.

Rojas, V. G., Loeb, S. C. and O'Keefe, J. M. (2018), 'False-positive occupancy models produce less-biased occupancy estimates for a rare and elusive bat species', *Journal of Mammalogy* **100**(1), 212–222.

Royle, J. A. and Link, W. A. (2006), 'Generalized site occupancy models allowing for false positive and false negative errors', *Ecology* **87**(4), 835–841.

Silvis, A., Perry, R. W. and Ford, W. (2016), Relationships of three species of White-nose syndrome-impacted bats to forest condition and management, General Technical Report SRS-214, US Forest Service Southern Research Station, Asheville, NC.

Stan Development Team (2020), 'RStan: the R interface to Stan'. R package version 2.21.2.

URL: <http://mc-stan.org/>

- Stratton, C., Irvine, K. M., Banner, K. M., Wright, W. J., Lausen, C. and Rae, J. (in press 2022), ‘Coupling validation effort with *in situ* bioacoustic data improves estimating relative activity and occupancy for multiple species with cross-species misclassifications’, *Methods in Ecology and Evolution* .
- Sugai, L. S. M., Silva, T. S. F., Ribeiro, José Wagner, J. and Llusia, D. (2018), ‘Terrestrial passive acoustic monitoring: Review and perspectives’, *BioScience* **69**(1), 15–25.
- Udell, B., Straw, B., Cheng, T., Enns, K., Winfred, F., Gotthold, B., Irvine, K., Lausen, C., Loeb, S., Reichard, J., Rodhouse, T., Smith, D., Stratton, C., Thogmartin, W., Wiens, A. and Reichert, B. (2022), ‘Status and trends of North American bats summer occupancy analysis 2010-2019 data release: U.S. Geological Survey data release’.
- USFWS (2020), Range-wide Indiana bat summer survey guidelines, Technical report, U.S. Fish and Wildlife Service Region 3, St. Paul, MN.
- USFWS and USGS (2019), Testing procedures, performance criteria and approval process for automated bat ID software programs associated with range-wide Indiana bat summer survey guidelines, Technical report, U.S. Fish and Wildlife Service Midwest Region, Bloomington, MN.
- Wade, P. R. (2000), ‘Bayesian methods in conservation biology’, *Conservation Biology* **14**, 1308–1316.
- Williams, P. J. and Hooten, M. B. (2016), ‘Combining statistical inference and decisions in ecology’, *Ecological Applications* **26**(6), 1930–1942.

Wright, W. J., Irvine, K., Alamber, E. and Litt, A. (2019), ‘Code release for bat data analyses in “Modelling misclassification in multi-species acoustic data when estimating occupancy and relative activity”: U.S. Geological Survey software release, <https://doi.org/10.5066/P9QK83LD>’.

Wright, W. J., Irvine, K. M., Almberg, E. S. and Litt, A. R. (2020), ‘Modelling misclassification in multi-species acoustic data when estimating occupancy and relative activity’, *Methods in Ecology and Evolution* **11**(1), 71–81.

Wright, W. J., Irvine, K. M., Rodhouse, T. and Litt, A. R. (2021), ‘Spatial gaussian processes improve multi-species occupancy models when range boundaries are uncertain and nonoverlapping’, *Ecology and Evolution* **11**(13), 8516–8527.

Tables

Scenario	No. Nights	No. Sites	Species	ϕ_k	ψ_k	λ_k	p_k
1	8, 16	55	Spp 1	(0.90, 0.35)	0.25	.3	0.26
			Spp 2	(0.10, 0.65)	0.75	10	0.999
2	8, 16	55	Spp 1	(0.65, 0.10)	0.25	.3	0.26
			Spp 2	(0.35, 0.90)	0.75	10	0.999
3	8, 16	55	Spp 1	(0.65, 0.40)	0.25	.3	0.26
			Spp 2	(0.35, 0.60)	0.75	10	0.999
4	8, 16	55	Spp 1	(0.90, 0.35)	0.50	.3	0.26
			Spp 2	(0.10, 0.65)	0.75	10	0.999
5	8, 16	55	Spp 1	(0.90, 0.35)	0.25	.7	0.50
			Spp 2	(0.10, 0.65)	0.75	10	0.999

Table 1: Parameter settings used to generate autoID datasets for 55 sites and 8 or 16 nightly surveys consistent with field knowledge and published work on *M. sodalis* (a rare species; denoted as “Spp 1”) and *M. lucifugus* (a common species; denoted as “Spp 2”). Simulated data assumes the information about auto-classifier performance identifying species is correct and constant for all sites and revisits. Note ϕ_k has the columns as the true species identity and the rows refer to the autoID label. For example, under Scenario 1, autoID labels to species 1 are correctly assigned 90% of the time and are contaminated by false positives or incorrectly labeled as species 1 35% of the time, on average. ψ_k is the probability species k occurs at a site, λ_k is the relative activity rate for species k during a nightly survey, and p_k is the probability at least one call file is recorded for species k during a nightly survey.

Figure Captions

Figure 1. Conceptual diagram representing our simulation study comparing four analytical approaches for determining local presence of rare bat species. The black pathways denote Bayesian model-based approaches that allow for both explanatory modeling and local decisions regarding species presence. The grey arrows represent the fact using the MLE-metric at the site-level (*MLESite*) to assess local species presence does not allow for estimating species-environment relationships more broadly (i.e., explanatory modeling). The stippled outlined boxes convey both the *Remove* and “*MLESite*” approach rely on the MLE-metric calculation at the visit- or site-level to adjust for species misclassifications. Dashed circles represent nodes that we investigated in the simulation study with accompanying factors (in call-out boxes) that we varied either through the data-generating process (relative activity, occurrence probability, and auto-classifier accuracy rates), design criteria (number of visits), or in the decision-rule applied to assign species presence at a site (cut-off values).

Figure 2. Comparison of average (thick line) and individual (thin lines) 95% posterior intervals from fitting Bayesian model-based approaches (*2SppCt*, *Remove*, or *Naive*). Rare species estimated probability of presence (ψ , top row), probability of detection (p , middle row), and relative activity (λ , bottom row) based on 16 visits and 55 sites under three different data-generating scenarios (corresponding column labels $[\psi = 0.25, \lambda = 0.3]$, $[\psi = 0.50, \lambda = 0.3]$, or $[\psi = 0.25, \lambda = 0.7]$). Horizontal black lines reflect the parameter values used for generating the datasets. Coverage of the data-generating parameter values indicated by the interval color (red = poor, purple = decent). There were a few iterations of the

789 simulation that resulted in MCMC samplers that did not converge ($R\text{-hat} > 1.1$) for the
790 *2SppCt* model and these iterations were excluded before average posterior intervals were
791 computed. Additional convergence details and scenarios included in Appendix S2: Figure
792 S1-S2.

793 Figure 3. Comparison of decision error rates for deciding rare species presence or absence
794 fitting the count detection model (labelled *2SppCt* darker green boxes) , *Remove* occupancy
795 model (labelled light green *Remove* boxes), or *MLESite* (mustard colored boxes). For *Remove*
796 and *2SppCt*, the posterior mean of the Z_{i1} -state $Prob(Z_{i1} = 1|y) > 0.75$ was used for species
797 absence, and $Prob(Z_{i1} = 1|y) > 0.25$ for species presence at a site i . The threshold-values
798 were selected to balance the (in)correct decision rates (See Appendix S2). The *MLESite*
799 used a p-value threshold of 0.10 in the decision-rule. Each dot represents the conditional
800 proportion of sites with a specific decision given the known true state of species present
801 ($Z = 1$) or absent ($Z = 0$) for each of 50 simulated datasets with 55 sites each. The columns
802 correspond to the true Z-state and each row is the species site-level decision. All scenario
803 results provided in Appendix S2, Figures S3-S5.

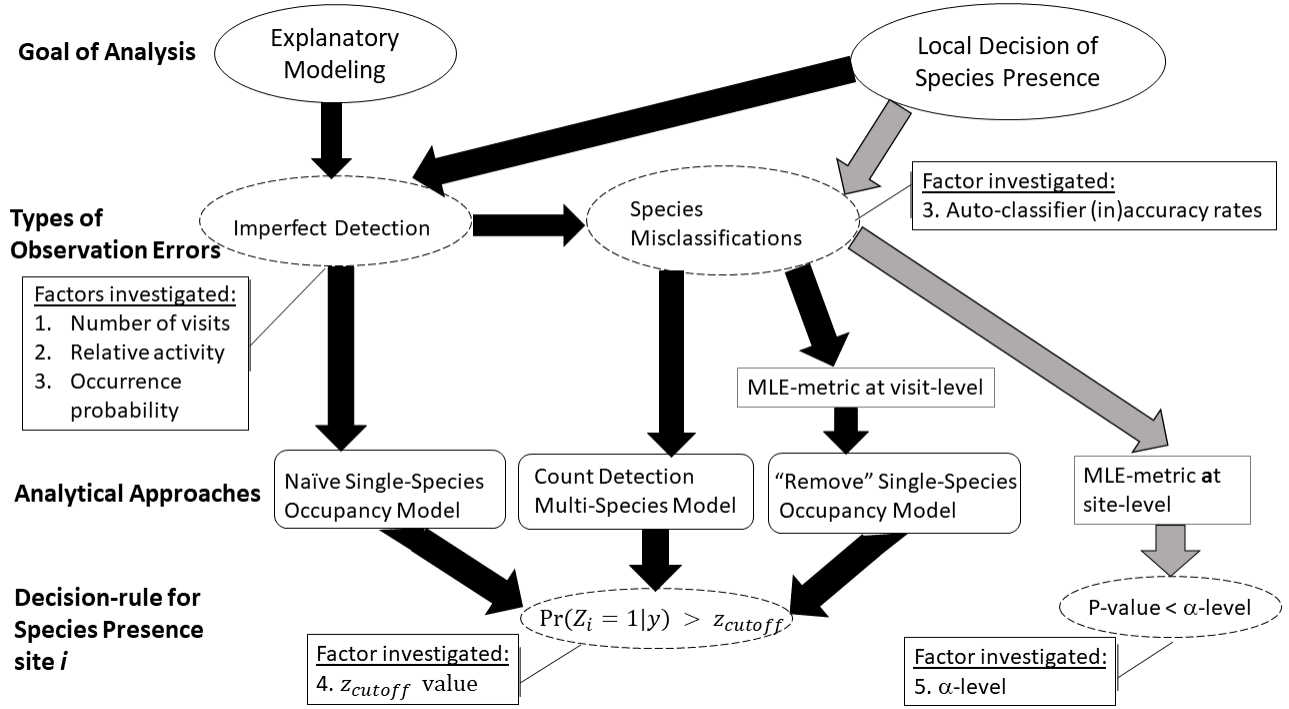


Figure 1: Conceptual diagram representing our simulation study comparing four analytical approaches for determining local presence of rare bat species. The black pathways denote Bayesian model-based approaches that allow for both explanatory modeling and local decisions regarding species presence. The grey arrows represent the fact using the MLE-metric at the site-level (*MLESite*) to assess local species presence does not allow for estimating species-environment relationships more broadly (i.e., explanatory modeling). The stippled outlined boxes convey both the *Remove* and “*MLESite*” approach rely on the MLE-metric calculation at the visit- or site-level to adjust for species misclassifications. Dashed circles represent nodes that we investigated in the simulation study with accompanying factors (in call-out boxes) that we varied either through the data-generating process (relative activity, occurrence probability, and auto-classifier accuracy rates), design criteria (number of visits), or in the decision-rule applied to assign species presence at a site (cut-off values).

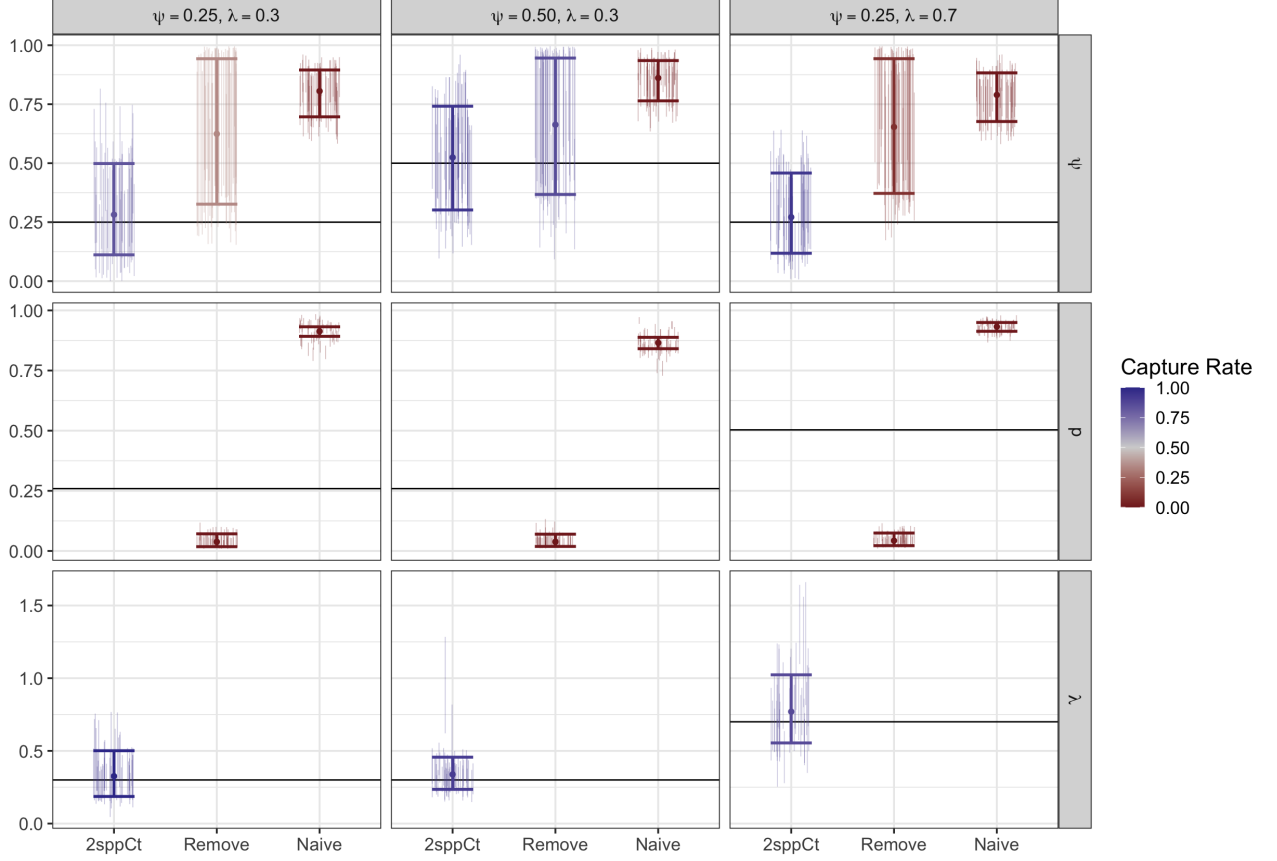


Figure 2: Comparison of average (thick line) and individual (thin lines) 95% posterior intervals from fitting Bayesian model-based approaches (*2SppCt*, *Remove*, or *Naive*). Rare species estimated probability of presence (ψ , top row), probability of detection (p , middle row), and relative activity (λ , bottom row) based on 16 visits and 55 sites under three different data-generating scenarios (corresponding column labels $[\psi = 0.25, \lambda = 0.3]$, $[\psi = 0.50, \lambda = 0.3]$, or $[\psi = 0.25, \lambda = 0.7]$). Horizontal black lines reflect the parameter values used for generating the datasets. Coverage of the data-generating parameter values indicated by the interval color (red = poor, purple = decent). There were a few iterations of the simulation that resulted in MCMC samplers that did not converge ($R\text{-hat} > 1.1$) for the *2SppCt* model and these iterations were excluded before average posterior intervals were computed. Additional convergence details and scenarios included in Appendix S2: Figure S1-S2.

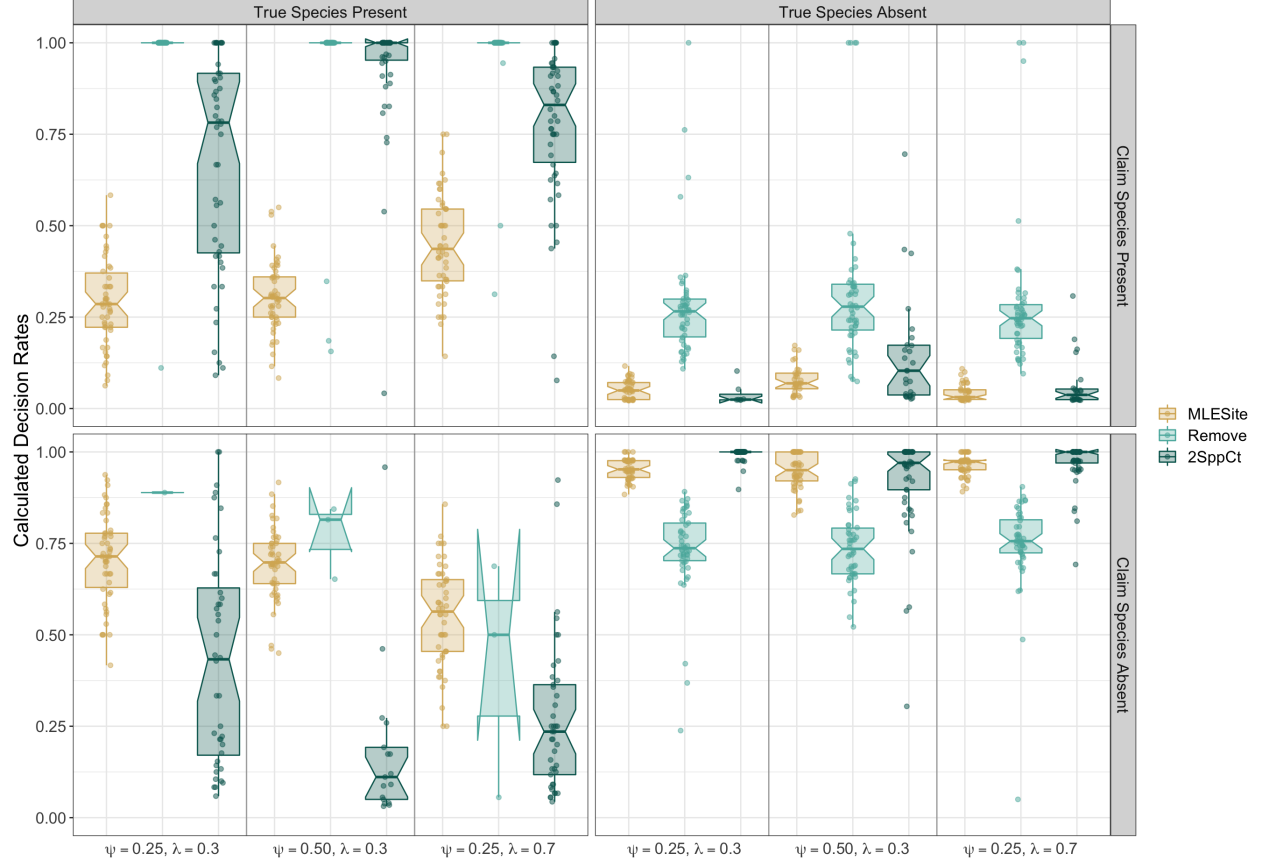


Figure 3: Comparison of decision error rates for deciding rare species presence or absence fitting the count detection model (labelled *2SppCt* darker green boxes), *Remove* occupancy model (labelled light green *Remove* boxes), or *MLESite* (mustard colored boxes). For *Remove* and *2SppCt*, the posterior mean of the Z_{i1} -state $Prob(Z_{i1} = 1|y) > 0.75$ was used for species absence, and $Prob(Z_{i1} = 1|y) > 0.25$ for species presence at a site i . The threshold-values were selected to balance the (in)correct decision rates (See Appendix S2). The *MLESite* used a p-value threshold of 0.10 in the decision-rule. Each dot represents the conditional proportion of sites with a specific decision given the known true state of species present ($Z = 1$) or absent ($Z = 0$) for each of 50 simulated datasets with 55 sites each. The columns correspond to the true Z-state and each row is the species site-level decision. All scenario results provided in Appendix S2, Figures S3-S5.

**Appendix S1. Statistical assessment on determining
local presence of rare bat species. K.M. Irvine, K.M.
Banner, C. Stratton, W.M. Ford, and B. Reichert. *Eco-
sphere***

Statistical Assumptions

An important consideration when fitting any statistical model is the underlying assumptions. If empirical data severely violate all or some of the assumptions, conclusions could be misleading. We made the following set of assumptions regarding bat acoustic data for our simulation study, and to show the connection between the MLE-metric p-value (Sec. 2.2; Britzke et al., 2002) and the recent multi-species count detection model developed for bioacoustic data (Sec. 2.3 Wright et al., 2020; Stratton et al., in press 2022).

The MLE-metric requires the following set of assumptions (A1 through A5) (Britzke et al., 2002).

A1: The assemblage of bat species available for detection at a location or site is known.

A2: An open-source machine learning algorithm or proprietary software is used to assign a species identity or label (hereafter, “autoID”) to a high frequency recording (sequence of echolocation pulses). The (in)accuracy rates of the automatic classification process are known and uniform among the available species communities or sites.

A3: A false positive is a result of a true positive from another species (a recording was

assigned the wrong species label).

A4: Identified recordings or observations are independent among species, nights, and sites.

A5: The data available for analysis are the number or count of recordings per species (autoIDs) for each site and recording interval (e.g., a single night or aggregated over many nights). Under field conditions, the automated classification process could result in a large number of recordings without a single-species label (e.g., a frequency group or “noID”). We assume those observations would be discarded prior to analysis, thereby reducing the total number of recordings to just those identified to a single species.

Mathematical connection between count detection posterior probability latent Z -state and “MLE” p-value

We show that the MLE-metric or the p-value from a likelihood ratio test for the null hypothesis of species absence versus the alternative hypothesis of species presence in the MLE calculation (Sec. 2.2) is related to the posterior probability species k is present at site i ($Prob[Z_{ik} = 0|y]$ or $Prob[Z_{ik} = 1|y]$) in the count detection model. To simplify notation, we consider 2 species ($K = 2$) and assume both ψ and λ do not vary among sites i or nights j (Eqn. 2 and 3 main text). However, the original Wright et al. (2020) multi-species count detection model was developed for more than two species and was applied to an acoustic dataset from Montana with seven species and an other category for the remaining species. Another recent application applied the same model to a bat acoustic dataset from British Columbia, Canada composed of 17 species (Stratton et al., in press 2022). For a direct comparison to the MLE-metric calculation by Britzke et al. (2002) we assume the data

844 available are the autoID counts per species (n_k) and the classifier (in)accuracy rates (ϕ) are
 845 known. For clarity, we let $\phi_{kk'}$ denote the probability that a call belonging to species k' is
 846 classified as species k , as in Britzke et al. (2002).

847 Under the count detection model, the sum of autoIDs for both species 1 (n_{i1}) and species
 848 2 (n_{i2}) are conditional on the latent Z-states (Z_{i1}, Z_{i2});

$$[n_1|(Z_{i1}, Z_{i2}, \lambda_1, \lambda_2)] = Poisson[Z_{i1}\lambda_1\phi_{11} + Z_{i2}\lambda_2\phi_{12}] \quad (4)$$

849 and

$$[n_2|(Z_{i1}, Z_{i2}, \lambda_1, \lambda_2)] = Poisson[Z_{i2}\lambda_2\phi_{22} + Z_{i1}\lambda_1\phi_{21}]. \quad (5)$$

850 Because the autoID counts are considered independent Poisson distributions, we can write
 851 the joint distribution for the autoIDs (n_{i1}, n_{i2}) as the product of a multinomial and Poisson
 852 distribution from statistical probability theory.

853 The multinomial is as follows conditional on the total number of recordings detected and
 854 identified at a site i ($N_i = n_{i1} + n_{i2}$) and the latent Z-states for a site i

$$[(n_{i1}, n_{i2})|(N_i, Z_{i1}, Z_{i2})] = Multinomial[N_i, \boldsymbol{\pi}_i], \quad (6)$$

855 where

$$\pi_{i1} = \frac{Z_{i1}\lambda_1\phi_{11} + Z_{i2}\lambda_2\phi_{12}}{Z_{i1}\lambda_1\phi_{11} + Z_{i2}\lambda_2\phi_{12} + Z_{i1}\lambda_1\phi_{21} + Z_{i2}\lambda_2\phi_{22}} \quad (7)$$

856 and

$$\pi_{i2} = \frac{Z_{i1}\lambda_1\phi_{21} + Z_{i2}\lambda_2\phi_{22}}{Z_{i1}\lambda_1\phi_{11} + Z_{i2}\lambda_2\phi_{12} + Z_{i1}\lambda_1\phi_{21} + Z_{i2}\lambda_2\phi_{22}}. \quad (8)$$

Both multinomial probabilities in Equation 7 and 8 simplify because $\phi_{11} + \phi_{21} = 1$ and $\phi_{12} + \phi_{22} = 1$ to the following,

$$\pi_{i1} = \left[\frac{Z_{i1}\lambda_1}{Z_{i1}\lambda_1 + Z_{i2}\lambda_2} \right] \phi_{11} + \left[\frac{Z_{i2}\lambda_2}{Z_{i1}\lambda_1 + Z_{i2}\lambda_2} \right] \phi_{12} \quad (9)$$

and

$$\pi_{i2} = \left[\frac{Z_{i2}\lambda_2}{Z_{i1}\lambda_1 + Z_{i2}\lambda_2} \right] \phi_{22} + \left[\frac{Z_{i1}\lambda_1}{Z_{i1}\lambda_1 + Z_{i2}\lambda_2} \right] \phi_{21}. \quad (10)$$

The distribution for the total number of recordings N_i conditional on the latent Z-states follows a Poisson (by the reproductive properties of independent Poisson distributions), as follows

$$\begin{aligned} [N_i|(Z_{i1}, Z_{i2})] &= \text{Poisson}(Z_{i1}\lambda_1\phi_{11} + Z_{i2}\lambda_2\phi_{12} + Z_{i2}\lambda_2\phi_{22} + Z_{i1}\lambda_1\phi_{21}) \\ &= \text{Poisson}(Z_{i1}\lambda_1 + Z_{i2}\lambda_2) \end{aligned} \quad (11)$$

and then

$$[Z_{ik}] = \text{Bernoulli}(\psi_k) \quad (12)$$

because of the independence assumptions for the latent Z-states.

The full likelihood is then the product over i and the Multinomial (Equation 6), Poisson (Equation 11), and then the Bernoulli (Equation 12), as follows

$$\prod_i [(n_{i1}, n_{i2})|(N_i, Z_{i1}, Z_{i2})][N_i|(Z_{i1}, Z_{i2})][Z_{i1}][Z_{i2}]. \quad (13)$$

We draw the connection to the MLE calculation by equating both the statistical assump-

868 tions (previous section) and the bottom-level of the Bayesian hierarchical model (Equation
 869 6 with simplified probabilities Equation 9 and 10) to the likelihood specified for the MLE
 870 calculation (Britzke et al., 2002).

871 In the two species case, Britzke et al. (2002) assumed a multinomial distribution for the
 872 autoIDs as follows (suppressing the additional sub-scripts of i and j for clarity),

$$\text{Multinomial} [N, (\phi_{11}\theta_1^B + \phi_{12}\theta_2^B, \phi_{21}\theta_1^B + \phi_{22}\theta_2^B)] . \quad (14)$$

873 The crux of our argument is that the same quantities (n_1, n_2) under the count detection
 874 model were modeled as,

$$\text{Multinomial} \left[N, \left(\phi_{11} \frac{Z_1 \lambda_1}{Z_1 \lambda_1 + Z_2 \lambda_2} + \phi_{12} \frac{Z_2 \lambda_2}{Z_1 \lambda_1 + Z_2 \lambda_2}, \phi_{21} \frac{Z_1 \lambda_1}{Z_1 \lambda_1 + Z_2 \lambda_2} + \phi_{22} \frac{Z_2 \lambda_2}{Z_1 \lambda_1 + Z_2 \lambda_2} \right) \right] . \quad (15)$$

875 Then by matching up the parameters in Equation 14 and 15, we have $\theta_k^B = \frac{Z_k \lambda_k}{\sum_k Z_k \lambda_k}$.
 876 The relative frequency for a species within the community (θ^B) was decomposed within
 877 the hierarchical count detection model into the latent state of interest (Z_k) and the relative
 878 activity for a species (λ_k). Therefore, a test for $\theta_k^B = 0$ in Eqn 14 was similar to $Pr(Z_k = 0|y)$
 879 in Eqn 15 because if $Z_k = 0$, then $\frac{Z_k \lambda_k}{\sum_k Z_k \lambda_k} = 0$. An important distinction is that with the
 880 multi-species count detection model we account for potential false negatives at a site-level
 881 [non-detections] and variable species relative activity among sites and revisits.